Blackwell Publishing

Cladistics 24 (2008) 62–73

Cladistics

10.1111/j.1096-0031.2007.00169.x

The unfalsifiability of cladograms and its consequences

L. Vogt*

Department of Organismic & Evolutionary Biology, Harvard University, Biolabs 1113, 16 Divinity Avenue, Cambridge, MA 20138, USA

Accepted 28 March 2007

Abstract

Popper's falsificationism provides the normative reference system in recent discussions regarding theory and methodology of systematics. According to Popper, the falsifiability of a hypothesis represents a necessary precondition for its corroborability. It is shown that cladograms, independent of "strict", "methodological" or "sophisticated" falsification, are not falsifiable in principle. No present observation is prohibited by any tree hypothesis and, thus, no Popperian test of cladograms exists. It is shown that the congruence test, which is commonly said to represent a Popperian test of cladograms, instead tests sets of apomorphy hypotheses. Three different strategies that have been proposed to circumvent this problem are discussed and refuted: (1) referring to Popper's convention to renounce *ad hoc* maneuvers; (2) referring to Popper's treatment of probability hypotheses; and (3) decoupling corroboration from falsification. As a consequence, within a Popperian framework the unfalsifiability of cladograms implies that cladograms cannot explain any present day observation and, thus, represent metaphysical hypotheses.

However, Popper's falsificationism has been criticized and questioned by many philosophers before and it seems to be about time that phylogeneticists develop their own philosophy of phylogenetics that meets their specific requirements of a historical science that is not seeking for universal laws and regularities, but instead reconstructing particular historical events.

© The Willi Hennig Society 2007.

The theory of biological classification in general and the methods of phylogenetics in particular have a long history of controversial discussions, resulting in a continuous improvement of theoretical concepts and analytical methods—distinction between ancestral and derived character states, application of algorithms and computer techniques for tree reconstruction, making explicit what evidence has been used by accounting for it with a character matrix, increasing availability of molecular data, to mention a few (see also O'Hara, 1997). Therewith, philosophical ideas often had a significant impact. Outstanding is the strong influence that Popper's falsificationist approach had on these discussions during the last three decades (for an overview see Helfenbein and DeSalle, 2005).

Interestingly, Popper's falsificationism provides the normative reference system that has not been questioned seriously by biologists so far, although many philosophers have criticized it and even questioned its foundations (e.g., Salmon, 1967, 1968, 1998; Lakatos, 1968, 1970; Kuhn, 1970; Putnam, 1974; Grunbaum, 1976; Mackie, 1985; Sober, 1988, 2000; Howson and Urbach, 1989; Earman and Salmon, 1992; McGuire, 1992; Stamos, 1996; Andersson, 1998; Schurz, 1998; Franklin, 2001; Spohn, 2001). This is even more surprising when considering the divergent and sometimes contradictory points of view phylogeneticists take in with respect to the implications from falsificationism for phylogenetic methodology (e.g., Farris, 1986, 1995; Kluge, 1997a, 2001a; de Queiroz and Poe, 2001, 2003; Faith and Trueman, 2001).

The evaluation of the consequences and the premises of a falsificationist approach to phylogenetic research represents an extensive and still ongoing discussion. Early discussions were much concerned with competing types of classifications and whether the theory of evolution and biological classification meet Popper's demarcation criterion for empirical science (e.g., Bock, 1969, 1973; Wiley, 1975; Engelmann and Wiley, 1977; Kitts, 1977; Platnick

*Corresponding author:
E-mail address: lars.vogt@zoosyst-berlin.de

© The Willi Hennig Society 2007

and Gaffney, 1977, 1978; Cracraft, 1978; Settle, 1979; Hull, 1980). Even by then different and partially contradictory positions have been advocated, all arguing their respective case with Popper. This holds true for recent discussions as well, although these discussions rather focus on the (philosophical) justification of the choice of the best analytical method for phylogenetic tree inference (Farris, 1986, 1995; de Queiroz and Poe, 2001, 2003; Farris et al., 2001). In particular, the proponents of cladistic parsimony extensively referred to falsificationism in order to defend the claim of superiority of their method and to criticize other methods such as likelihood as verificationist (e.g., Farris, 1983; Kluge, 1997a,b; Siddall and Kluge, 1997). By applying a specific interpretation of Popper's concept of corroboration to phylogenetics they claim (among other things) that the most corroborated tree is the tree with minimum length because it requires the least amount of homoplasies, which at their turn are understood to represent Popperian *ad hoc* hypotheses whose amount has to be minimized (e.g., Farris, 1983, 2000; Kluge, 1997a,b, 2001a, 2002).

As Popperian falsificationism takes in such a central role within the respective theoretical discussions I will take in the Popperian point of view throughout this paper. Therewith, I will not discuss the drawbacks or advantages of falsificationism itself and the practical and theoretical problems that are involved with its actual application to science in general and historical sciences in particular. Instead, I will infer and discuss the consequences of consistently applying falsificationism to phylogenetics. Therefore, this paper represents a thought-experiment: Let's assume Popperian falsificationism is really the only reasonable and justifiable way to do empirical research—What would this imply for phylogenetics?

In the following I will demonstrate that with respect to the common conditions of phylogenetic analyses phylogenetic tree hypotheses (i.e., cladograms) are not falsifiable in principle and, thus, cannot gain corroboration. As tree hypotheses cannot gain corroboration and as each tree transcends the observable, there is no strictly Popperian approach for justifying the choice of the best tree in light of present empirical evidence. I will argue that the three different strategies that have been advocated to circumvent this problem all are not consistent with falsificationism. Consequently, *sensu* Popper's demarcation criterion of falsifiability, seeking phylogenetic trees does not represent a scientific endeavor and cladograms represent no scientific but metaphysical hypotheses.

Popper's falsificationism

By referring to Hume's problem of induction (Hume, 1748/1993), Popper concluded that empirical sciences could only be demarcated from metaphysics if their methodology of scientific inquiry exclusively rests on deductive logic without requiring induction at all (Popper, 1935/1994, p. 14f.). The reason for the restriction to deduction is because deduction is the only truth-keeping type of logical reasoning and is therefore free of the problems that Hume identified regarding induction.

According to Popper, scientific hypotheses have to be challenged constantly by experience. This can be accomplished by deducing *necessary* predictions from the hypothesis and subsequently testing the hypothesis by comparing the predictions with actual observations. These predictions usually have the form of prohibitions. Therewith, the so-called hypothetico-deductive setting plays a central role within Popper's approach, consisting of a temporarily accepted relevant background knowledge b, an empirical hypothesis h, and basic statements e (e.g., Popper, 1935/1994, p. 31ff; Popper, 1983, p. 236ff). The hypothetico-deductive setting provides the basis on which the empirical tests can be conducted. Basic statements, which are observational statements or empirical hypotheses of lower universality that have been corroborated by tests against empirical statements at a lower level of investigation (Rieppel and Kearney, 2002), serve as falsifiers in case their existence is prohibited by the hypothesis in question (Popper, 1935/1994, p. 66ff, 77ff; Popper, 1983, p. 217ff; see also Grant and Kluge, 2003). Thus, falsification necessarily requires the applicability of deduction (Grant and Kluge, 2003), following the schema of *modus tollens*, *modus ponens* and the use of *transposition*.

Modus tollens argument:
If (h & b), then *non-e*.
non-e is false.
Therefore, (h & b) is false.

Modus ponens argument:
If (h & b), then *non-e*.
(h & b).
Therefore, *non-e*.

Popper's use of *modus tollens* (in combination with *modus ponens* and transposition):

If (h & b), then *non-e*.	(premise – material implication)
If *non-e* false, then (h & b) false.	(derived by transposition)
non-e false.	(premise)
Therefore, (h & b) is false.	(derived by *modus ponens*)

From the first premise of *modus tollens* follows that *non-e* is a *necessary condition* of the conjunction of hypothesis and background knowledge—the hypothesis prohibits the occurrence of e. Therefore, in case e is given it follows that the conjunction of hypothesis and background knowledge is necessarily false.

To be a Popperian test, potential falsifiers have to be *deducible* from the hypothesis to guarantee the hypothesis' falsifiability—hypotheses must prohibit specific

observations that are not prohibited by the background knowledge, in order to be called scientific hypotheses in the Popperian sense. As a consequence, following *modus tollens*, the hypothesis would be falsified by the occurrence of contradicting empirical evidence given that the evidence and background knowledge were true. However, in case an actual observation contradicts the prediction "*non-e*", we can only conclude that the hypothetico-deductive setting as a whole is inconsistent and that at least one of the three components—the background knowledge including initial conditions and auxiliary hypotheses, the falsifying basic statement, or the hypothesis—necessarily has to be false. The hypothesis could be considered to be falsified *only* under the assumption that the background knowledge and the basic statement are true, which cannot be known in principle (Duhem–Quine Thesis; Duhem, 1906; Quine, 1951). This represents a well known limitation of falsification and results from the logical relations of the components of the hypothetico-deductive setting: A hypothesis on its own cannot be falsified in principle as its consequences/predictions typically rest on background knowledge assumptions and, thus, depend on them. This is the reason for the use of the hypothetico-deductive setting in hypothesis testing. According to the Duhem–Quine Thesis, only the hypothetico-deductive setting *as a whole* can be falsified, without knowing which of the three components of the hypothetico-deductive setting is responsible for falsification. Ignoring this limitation of falsificationism would be to commit fallacy of naïve "strict falsification".

This limitation, however, does not affect Popper's criterion of falsifiability: a hypothesis is *falsifiable* in principle if it prohibits observations that are not prohibited by the background knowledge, although an actual *falsification* of the hypothesis in the strict sense is not possible as, according to Popper, basic statements and background knowledge are not verifiable in principle (e.g., Popper, 1983, p. xxii). Thus, Popper's falsifiability refers to the *logical* and not the *epistemological* status of the hypothesis (sometimes also referred to as *methodological* or *logical falsifiability* as opposed to *strict* or *empirical falsification*).

Popper developed the concept of logical probability for his falsificationist approach—the more a hypothesis prohibits, the higher is its amount of potential falsifiers, its empirical content, and the more logically *im*probable is the hypothesis (Popper, 1935/1994, pp. 78, 84). Moreover, as testability of a hypothesis directly depends on the amount of potential falsifiers of the hypothesis and as potential explanatory power depends on the testability of the hypothesis, according to Popper, empirical content, degree of falsifiability, degree of testability, potential explanatory power, and logical improbability of a hypothesis all correlate (e.g., Popper, 1935/1994, pp. 78, 84). Consequently, the more improbable a hypothesis the more it potentially explains.

The severity of a Popperian test determines the degree of corroboration a hypothesis gains when passing the test successfully. Therewith, degree of corroboration designates the degree of acceptability of a hypothesis in the light of our present experience (e.g., Popper, 1935/1994, p. 77ff; Popper, 1983, p. 217ff). According to Popper (e.g., Popper, 1935/1994, p. 77ff; Popper, 1983, p. 220ff, p. 231), the degree of severity of a test depends on the amount of "different" falsifiers that the test *potentially* accredits.

The two concepts, falsification and corroboration, are linked to one another via the ideas of degree of falsifiability and degree of testability of a hypothesis that determine the upper limit of the possible degree of corroboration a hypothesis can potentially gain through tests; in other words its corroborability. The degree of explanatory power of a hypothesis always correlates with its current degree of corroboration and provides the basis for choosing among many possible hypotheses the best hypothesis that is presently available in the light of the results of empirical tests. Thus, in Popperian falsificationism the inference to the best explanation for given empirical evidence breaks down into inferring the hypothesis with the highest degree of corroboration among all possible alternatives. As corroboration can only be gained by performing potentially successful attempts of falsifying a hypothesis, only (methodologically) falsifiable hypotheses can qualify for the best explanation.

Thus, according to Popper an observation is explained by a hypothesis in case the hypothesis successfully passed the most severe tests and the corresponding observational statement is logically, i.e., deductively, entailed in the conjunction of hypothesis and assumed background knowledge. This model of explanation is known by many names, including the covering law model, the Hempel–Oppenheim model, the Popper–Hempel model, or the deductive-nomological model of explanation (Niiniluoto, 1995).

Distinguishing falsifiability and falsification

Important for understanding Popper's criterion of falsifiability is his distinction of falsification and falsifiability. According to Popper, *falsifiability* refers to the logical status of a statement, whereas *falsification* refers to the practical procedure of empirically testing a hypothesis—he often uses the term logical falsifiability and definitive, conclusive, demonstrable, or empirical falsification to distinguish them (e.g., Popper, 1983, p. xxii). Popper thereby outspokenly admits that while the former is possible, the latter is practically impossible due to the fallibility of both our observations and the assumed background knowledge and the logical dependence of the hypothesis and the background knowledge within the hypothetico-deductive setting (see Duhem–Quine Thesis above).

Both falsification and corroboration require falsifiability, while falsifiability does not necessarily require the possibility of actual falsification or corroboration. Thus, before one investigates the possibilities of empirically testing a hypothesis one first has to ascertain its falsifiability. As a consequence, when discussing the falsifiability of phylogenetic hypotheses for a start it is not necessarily relevant whether they actually can be falsified in practice.

According to Popper, a hypothesis is only falsifiable in case its empirical content, i.e., the class of its potential falsifiers, is not empty (Popper, 1935/1994, pp. 78, 84). Otherwise, *modus tollens* is not applicable. Thus, a falsifiable hypothesis has to prohibit observational statements that are not prohibited by the background knowledge. In order to guarantee the falsifiability of a specific phylogenetic hypothesis one has to identify at least one potentially falsifying observational statement, i.e., a statement that is only prohibited by the hypothesis but not by the assumed background knowledge. This is an indispensable logical prerequisite—a necessary precondition for the possibility of empirical tests in phylogenetics and thus precedes any consideration and discussion of the application of Popper's concept of corroboration in phylogenetics.

Are cladograms falsifiable?

What could serve as a basis for an empirical test and which type of phylogenetic hypothesis is falsifiable has already been discussed in the 1970s. For instance, Bock (1973) proposes that hypotheses of sister group relationships as well as hypotheses of ancestral-descendent relationships are testable in principle. However, proper methods of testing still had to be developed. Bock mentions the consistency test of Wilson (1965) as an example for a possible test. By a procedure of disproving more and more hypotheses of phylogenetic relationships and thereby eliminating more and more alternative explanations until only one is left, Bock (1973) predicts that the application of the Popperian philosophy will lead to a condition where relative weights of characters will play no significant role anymore. From our present day perspective, however, Bock's appraisal of the future impact of Popper's approach has to be judged to be rather enthusiastic.

In recent discussions opinions about potential mechanisms of falsification within phylogenetics are contradictory. This holds true particularly for the aspect of testing. Kluge (1997a, 2002) states that synapomorphies can be used to falsify cladograms. In the same paper, however, Kluge (1997a, p. 86) concludes that in the case of cladograms deduction would be "impossible, because a cladogram is logically consistent with all synapomorphy distributions, congruent and incongruent", which renders cladograms unfalsifiable under these circumstances.

Neither such background knowledge as for instance "*descent with modification*", nor any specific tree hypothesis prohibits the occurrence of convergent evolution. This allows for both apomorphy (throughout this paper, I will use the term "apomorphy" to mean *structural sameness due to shared common origin* and not simply *observational similarity*) and homoplasy as possible explanations for the sameness of character states and their distribution patterns. A given tree hypothesis is logically congruent with any specific observable evidence of character state distribution. In other words, a given tree, in combination with decent with modification as background knowledge, *does not prohibit* any specific character state distribution pattern (Farris, 1983; Sober, 1983). As there is no deductive link between any tree hypothesis and any specific character state distribution there exists no direct empirical test of hypotheses of monophyly (i.e., clades) *sensu* Popper (Sober, 1988; Rieppel, 2003)—one cannot think of any observation, which, in case it would represent a true statement, would allow to conclude the falsity of a clade or a given cladogram through *modus tollens*. As a consequence, and this is independent of the distinction of naïve "strict falsification" and "methodological" or "sophisticated falsification", hypotheses of phylogenetic relationships are *not falsifiable* in principle as cladograms are not directly testable in the Popperian sense—though they are fallible of course (see also Hull, 1983; Sober, 1983; Rieppel, 2003; contradicting: Bock, 1973; Cracraft, 1978; Farris, 1983, 2000; Kluge, 1997a,b, 2003; Farris et al., 2001; but see Kluge, 1999; Grant and Kluge, 2003).

Testing cladograms with congruence

The *criterion of congruence* is commonly referred to as a test of cladograms that is performed during phylogenetic tree analysis and that requires a set of character hypotheses as given (Patterson, 1982, 1988; de Pinna, 1991; Kluge, 1997a; see also criterion of coincidence, Wagner, 1986; consistency test, Wilson, 1965). Thus, it is not surprising that some authors try to conceive this test as a Popperian test of cladograms (e.g., Farris, 1983, 2000; Bryant, 1992; Kluge, 1997a, 1999; Grant and Kluge, 2003), although it was not originally introduced in a Popperian context (see Kearney and Rieppel, 2006).

Authors that conceive the congruence test as a Popperian test usually assume a hypothetico-deductive setting consisting of a phylogenetic tree as the hypothesis to be tested, descent with modification as assumed background knowledge, and a set of synapomorphy hypotheses as basic statements. As has been shown

above, in case "synapomorphy" is understood as a purely observational statement of the distribution pattern of same traits shared by different taxa, the congruence test does not represent a Popperian test—at least on the basis of the above-mentioned hypothetico-deductive setting (contradicting Kluge, 1997a; Grant and Kluge, 2003). On the other hand, in case "synapomorphy" is understood to represent a distribution pattern of same traits that originated from a common ancestor, "synapomorphy" cannot serve as Popperian basic statement. Statements about distribution patterns of character states (i.e., putative apomorphies) do not represent statements that directly refer to observations as they represent hypotheses of homology. Hypotheses of homology go beyond observational statements as they imply the assumption of a common causal origin to the observable sameness of structures of different organisms. The corresponding distribution patterns of homologous states, as they themselves represent hypotheses, can only serve as basic statements in the congruence test in case they gained corroboration by already having passed a Popperian test against observational statements beforehand (Vogt, 2002, 2004a; Rieppel and Kearney, 2002). If this is not the case, the congruence test merely represents a purely formal test of the logical relation of statements without any empirical ground or correspondence (Vogt, 2002, 2004a; see also Rieppel, 2005; Kearney and Rieppel, 2006).

Irrespective of this result it can be questioned whether the congruence test tests phylogenetic tree hypotheses at all. In order to better understand the actual test mechanism of the congruence test one has to consider the asymmetrical relation between the concept of apomorphy and the concept of monophyly. In case a specific hypothesis of apomorphy could be falsified one could *not* deduce that the corresponding hypothesis of monophyly is necessarily falsified as well since another organismic structure may exist that represents an apomorphy in its own right and that exhibits the same distribution pattern therewith coding for the same specific monophylum. This is due to the fact that one cannot conclude from the falsification of the presence of particular evidence for the falsification of the presence of any evidence at all for the corresponding hypothesis of monophyly (analog to "absence of evidence is not evidence of absence"). On the other hand, *if* one could falsify a hypothesis of monophyly one could logically conclude that all possible corresponding hypotheses of apomorphy are *necessarily* falsified as well. Following this asymmetrical relation one can say that falsifying an apomorphy hypothesis does not at the same time falsify its corresponding monophyly hypothesis, but if a specific monophyly hypothesis or a combination of such is prohibited, the corresponding apomorphy hypothesis or a combination of them is *necessarily* prohibited as well.

This point is very important for the understanding of the mechanism of the congruence test: From the assumption of a bifurcating mode of speciation follows necessarily that clades (i.e., monophyla) are prohibited to overlap one another. Not all possible monophyletic groups can exist simultaneously. This allows for the testing of *sets of hypotheses of apomorphy* on Popperian grounds: during the congruence test sets of distribution patterns of hypotheses of apomorphy are compared against *all* theoretically possible *sets of congruent* monophyletic groupings of the corresponding operational taxonomic units. This test can only be performed if specific sets of monophyla are prohibited a priori—those that are incongruent, i.e., that overlap each other. From the a priori exclusion of all the sets of monophyla that contradict the necessity of an encaptic hierarchy, the falsification of the corresponding *sets* of hypotheses of apomorphy is deduced. This provides the basis for the congruence test. It would be *logically circular* to deduce the possibility to falsify specific hypotheses of monophyly from the congruence test as the test can only function by assuming that specific *sets of monophyly* are prohibited *as sets* beforehand.

In case of incongruence we only know that not all of the hypotheses of apomorphy of the tested set can represent true apomorphies—at least one of them has to represent a homoplasy (Wiley, 1975; Kluge, 1997a; according to the Duhem–Quine Thesis we would even have to consider that our assumed background knowledge or the basic statements can be responsible for the incongruency). Unfortunately, the congruence test does not indicate which specific hypothesis of apomorphy has been falsified—it only tells us that the *set as a whole is incongruent*. This is a consequence of the structure of this test in which single hypotheses of apomorphy cannot be tested on congruence with some *external* parameter but only a *set* consisting of at least two such hypotheses.

As a consequence, the congruence test as a Popperian test can provide corroboration only to *sets* of hypotheses of apomorphy and *not* hypotheses of monophyly or phylogenetic trees (contradicting Kluge, 1997a, 2002; Farris et al., 2001; Grant and Kluge, 2003). Therewith it is necessary that the basic statements of this test, i.e., the corresponding distribution patterns of the putative apomorphies of the tested set of hypotheses, successfully passed a Popperian test beforehand. Otherwise, the distribution patterns used in the congruence test do not qualify for Popperian basic statements and the congruence test is no Popperian test at all.

Cladograms and corroboration

Regarding Popperian falsificationism, the unfalsifiability of phylogenetic tree hypotheses has far reaching

consequences: the degree of testability and the degree of corroborability of clades necessarily equal zero. Clades and cladograms have no explanatory power at all. Following Popper, this implies that cladograms do not represent scientific hypotheses.

However, although some authors agree with the lack of a deductive link between observations and tree hypotheses, they still want to assign degrees of corroboration to tree hypotheses. For this reason they necessarily have to "re-interpret" Popper's concept of corroboration. In the following I will discuss three different strategies that follow this goal.

Minimizing ad hoc *hypotheses of homoplasy and the congruence test*

To provide a rational criterion for the choice of a presently preferred cladogram out of the plethora of alternative possible cladograms some cladistic phylogeneticists (e.g., Farris, 1983, 2000; Kluge, 1997a,b, 2001a, 2002) follow the strategy to shift the attention from the condition for corroborability, which is testability/falsifiability, to a methodological convention that Popper himself suggested *in addition* to his ideas of falsifiability, falsification, and corroboration—to renounce all *ad hoc* maneuvers (Popper, 1935/1994, p. 50f). Popper (1935/1994, p. 51) argues that only those *ad hoc* hypotheses should be accepted that do not decrease the falsifiability of the system. It has been argued that hypotheses of homoplasy represent phylogenetic *ad hoc* hypotheses in the Popperian sense. Therefore, choosing the minimum-step tree on grounds of parsimony as an optimality criterion would, in case one applies an equal weighting scheme, correspond with choosing the tree hypothesis that implies the minimum amount of *ad hoc* hypotheses and therewith the highest degree of corroboration (e.g., Farris, 1983; Kluge, 1997a,b). The idea is that tree hypotheses can gain corroboration in principle and that they do so not through the evidential weight of characters.

By referring to the congruence test and an asymmetrical relationship between (syn-)apomorphy, homoplasy and cladograms, Kluge (1997a) claims that a cladogram alone does not imply that congruent character states necessarily represent true apomorphies, but it does imply that incongruent character states are homoplasious. Although the author recognizes that the "test of a cladistic hypothesis cannot be a matter of deduction" (Kluge, 1997a, p. 86) and that the refutation of a specific cladogram, following the maxim of minimizing the requirement of homoplasies, must be understood as non-deductive (Kluge, 1997a, p. 87), he rates the lack of a deductive link irrelevant to testability in phylogenetic systematics as the "falsehood of a hypothesis can never be proven, even where deductive logic applies, because the falsifying observational proposition may itself be false". Kluge concludes that one should prefer "the hypothesis that requires the *ad hoc* dismissal of the fewest falsifiers" (Kluge, 1997a, p. 87). In other words, by implicitly referring to the general impossibility to verify observational statements (in part resulting from the assumption of the theory dependence of observation) and the Duhem–Quine Thesis, Kluge argues for the dismissal of the criterion of falsifiability in favor of the convention to renounce all *ad hoc* maneuvers, simply because we can never truly falsify a given hypothesis in the strict sense of naïve "strict falsification".

However, referring to the impossibility of strict practical falsification cannot compensate for the lack of falsifiability of cladograms, as falsifiability represents a necessary logical prerequisite for Popperian testability and corroboration, no matter whether one seeks for strict naïve, methodological, or sophisticated falsification (e.g., Popper, 1983, pp. 241, 245). Moreover, Popper himself acknowledged that the claim to renounce *ad hoc* maneuvers represents a methodological convention that is independent of falsification and corroboration (Popper, 1983, p. 133ff, p. 232; Popper 1935/1994, p. 16, 48ff, p. 105). As a consequence one cannot refer to the convention to renounce all *ad hoc* maneuvers as a compensation for the lack of falsifiability in order to guarantee the corroborability of tree hypotheses, as Popper's approach provides no logical relation between corroborability and the convention that would justify this compensation.

Do homoplasy hypotheses represent Popperian ad hoc *hypotheses?*

Besides the problem that tree hypotheses cannot gain corroboration in principle, Kluge's approach may seem right on the first glance as homoplasy hypotheses are not testable in principle and should be avoided within a falsificationist framework whenever possible. Kluge (2001b, p. 202) gives the following summary of Popper's convention to renounce all *ad hoc* maneuvers: "*Ad hoc* hypotheses are usually introduced *a posteriori* into a discussion for the sole purpose of saving a cherished hypothesis from the threat of disconfirming evidence or theoretical inconsistency".

Another principal problem of Kluge's application of Popper's claim of renouncing all *ad hoc* maneuvers in phylogenetics lies in the fact that hypotheses of homoplasy do not qualify as *ad hoc* hypotheses in the Popperian sense. Nobody seriously denies that, when "descent with modification" is assumed as background knowledge, there are in principle two (with plesiomorphy even three) alternative *possible types of* explanations to the empirical phenomenon of indistinguishable traits between representatives of different species. These alternatives can be specified a priori the phylogenetic analysis—after all, this is one of the reasons why cladograms are not falsifiable in principle. On the one

hand, one *possible* explanation for the sameness of traits in different organisms and species and their distribution pattern is that they represent an *apomorphy*, going back to a single transformation event in their last common ancestor. On the other hand they can represent the result of several independent transformation events, in which case they represent a *homoplasy*. Both are consistent with the background knowledge of "descent with modification" and both have the potential to explain every possible distribution pattern of identical structures. Therefore, these two interpretations represent true alternative putative explanations (de Queiroz and Poe, 2003; contradicting Kluge, 2001b) and with respect to falsification one has to evaluate, which is more corroborated. The principle *possibility* of homoplasies is known *before* the numerical tree inference and is a necessary consequence from "descent with modification" as background knowledge. Furthermore, one cannot decide only on the basis of the structure itself whether the structure represents an apomorphy, a plesiomorphy, or a convergence.

Thus, hypotheses of homoplasy do not represent *ad hoc* hypotheses of the type Popper is talking about, which are conceived *after* falsification takes place, with the sole purpose of saving the hypothesis in question. According to Popper *ad hoc* maneuvers represent *a posteriori* modifications of the hypothetico-deductive setting. Often, such a modification comes in the form of additional conditions and auxiliary hypotheses that either changes the hypothesis itself, the background knowledge, or the basic statements and therewith saving the hypothesis from the attempt of falsification. The *ad hoc* reinterpretation of hypotheses of apomorphy to hypotheses of homoplasy in case of a falsification of a *set* of hypotheses of apomorphy during the congruence test *does not save* the tested set of hypotheses: the *set* of hypotheses of apomorphy that did not pass the congruence test *remains falsified as a set*, no matter how many characters are subsequently "reinterpreted" as homoplasies. In fact, such a "reinterpretation" of one or more apomorphy hypotheses would represent a decision in favor of the alternative explanation of convergence or plesiomorphy *because* the alternative, a set of apomorphy hypotheses, has been falsified.

As a consequence, hypotheses of homoplasy cannot be interpreted as *ad hoc* hypotheses in the Popperian sense, and, thus, this procedure cannot be justified on Popperian grounds—neither in reference to testability nor to Popperian *ad hoc* maneuvers. Furthermore, referring to Popper's convention of renouncing *ad hoc* maneuvers does not represent a valid argument in favor of cladistic parsimony and against maximum likelihood approaches in phylogenetics either (contradicting, e.g., Farris, 1983, 2000; Kluge, 1997a,b, 2001a, 2002).

A good example for a true Popperian *ad hoc* maneuver would be dismissing the assumption of a bifurcating mode of speciation in case of incongruence. In this way one could "save" a set of incongruent apomorphy hypotheses from falsification during the congruence test by changing the background knowledge: without the assumption of a bifurcating mode of speciation monophyletic groups would be allowed to overlap each other and apomorphies would be allowed to be incongruent.

Referring to Popper's treatment of probabilistic hypotheses

A different strategy has been advocated by de Queiroz and Poe (2001, 2003). They claim that likelihood would not only be "consistent with Popper's concept of corroboration" but that it is, "as is evident in Popper's own writings, the foundation of Popper's concept [of corroboration]" (de Queiroz and Poe, 2001, p. 308). According to de Queiroz and Poe (2001, 2003), the tree with the highest likelihood is the most corroborated tree. However, Popper (1935/1994, p. 343) himself unequivocally states that, because of formal as well as intuitive reasons, an equation of degree of corroboration with probability as well as with likelihood would be absurd—the equation would lead to a logical contradiction (see also Kluge, 2001a). Besides that, there are other issues with the approach of de Queiroz and Poe. de Queiroz and Poe (2003, p. 362f) address the question of falsifiability of tree hypotheses by referring to Popper's treatment of probabilistic hypotheses (e.g., throwing coins, Popper, 1935/1994, p. 145), therewith claiming that phylogenetic hypotheses represent probabilistic hypotheses like Popper had in mind.

Popper argues that, although no evidence e can be found that could falsify a given probabilistic hypothesis h in the sense that $p(e,hb)$ equals zero (Popper, 1983, p. 242), for some evidence scientists are well able to decide whether a particular probabilistic hypothesis ought to be rejected as "practically falsified" (Popper, 1935/1994). By equating phylogenetic tree hypotheses with the probabilistic hypotheses specified by Popper, de Queiroz and Poe (2003) argue for the general possibility of assigning degrees of corroboration to cladograms in dependence of the likelihood values they obtain irrespective of the fact that no evidence can be specified that could methodologically falsify a given tree hypothesis.

Besides the fact that de Queiroz and Poe thereby refer to one of the rather disputed parts of Popper's approach—the testing of probabilistic hypotheses (e.g., Schroeder-Heister, 1998)—it is questionable that cladograms represent the kind of probabilistic hypotheses that Popper had in mind. Popper's probabilistic hypotheses are of the kind "$p(a,b) = r$"—a probability statement of the type "when b then a with a probability of r". Thus, the probability value is part of the probabilistic hypothesis itself and it is this value that is tested in

Popperian probabilistic hypothesis testing. Popper realized that the problem of such probabilistic hypotheses with respect to falsifiability is that one would need an infinite conjunction of basic statements in order to be able to refute or falsify them, which is practically impossible (Popper, 1935/1994, p. 144ff). In order to save the scientific status of such probabilistic hypotheses Popper introduced another methodological convention based on a decision as to whether the observed frequency of events that are employed as evidence in the test is representative for the infinite set of all past and future events that in theory would be required for falsification (Schroeder-Heister, 1998).

Unfortunately, although likelihood values can be assigned *to* phylogenetic tree hypotheses, tree hypotheses *themselves* do by no means represent probabilistic hypotheses of the kind Popper described. This is true because even an infinite number of characters would not provide the required logical basis for the falsifiability of tree hypotheses. That likelihood analyses result in the designation of likelihood values, which assign a probability to the evidence in the light of a given tree hypothesis and background knowledge, is irrelevant under these circumstances, as the likelihood values themselves are not subject to tests during tree inference. They rather represent the *result* of a likelihood tree inference and do not represent hypotheses that are tested. de Queiroz and Poe (2001, 2003) make the mistake to mix up the likelihood term $p(e,hb)$, with a given phylogenetic tree hypothesis h to be tested—but $p(e,hb) \neq h$. It would also be circular to understand the likelihood term as providing the probabilistic basis that guarantees the falsifiability of tree hypotheses and at the same time understand likelihood as providing a measure for the degree of corroboration of a tree resulting from a Popperian test.

Decoupling corroboration from falsification

Another strategy that has been propagated is to take some degree of goodness-of-fit as the relevant evidence for phylogenetic hypothesis testing. For instance, Cracraft and Helm-Bychowski (1991) advocate that independent data corroborate a tree when it is congruent with the most parsimonious or most likely tree that has been obtained from other data. In some respects, a similar strategy has been advocated and justified in reference to Popperian falsificationism by Faith (1991, 1992, 1999, 2004, 2006), Faith and Cranston (1992), and Faith and Trueman (2001). These authors introduced the argument that degree of corroboration of a phylogenetic hypothesis is primarily indicated by the improbability of data as fit-as-evidence, rather than some count of the number of times the tree hypothesis could not be "falsified" (Faith and Trueman, 2001, p. 332). As a consequence, instead of character state distributions they understand goodness-of-fit as the source for basic statements of Popperian hypothesis testing of phylogenetic tree hypotheses. Instead of emphasizing falsification and falsifiability, the authors demand that one should emphasize on Popper's concept of corroboration for which evidence does not necessarily involve any demonstrable falsification, but for which the improbability of the evidence in the absence of the hypothesis, i.e., $p(e,b)$, is highly relevant.

According to Faith and Trueman (2001) one could differentiate between a "falsification" and a "corroboration/severity" interpretation of Popperian philosophy, and the requirement of improbability of evidence in the absence of the hypothesis would represent a "key defining property for Popperian corroboration" (Faith, 2004, p. 2). Applied to phylogenetics, Faith and Trueman call their approach the "inclusive framework". In the "inclusive framework" the improbability of goodness-of-fit as evidence is understood to significantly contribute to the corroboration of phylogenetic tree hypotheses. Evidence e for a tree hypothesis h is typically given by a measure of the goodness-of-fit of the observed character data to the tree hypothesis (obtained, e.g., with parsimony or likelihood) while corroboration of a tree hypothesis is given by the improbability of that goodness-of-fit (Faith, 2004).

Faith and Trueman (2001) argue on grounds of Popper's concept of logical probability, from which follows that a low logical probability of a hypothesis implies high empirical content for the hypothesis and vice versa. Taking $p(e,h)$ as a statement implying degree of fit of hypothesis with evidence, it would follow that for good fit, $p(e,h)$ would be high and the content of e low given h. According to Faith and Trueman (2001), high content of the hypothesis is indirectly indicated by finding low $p(e,b)$ values for any evidence that follows from the hypothesis. As a consequence, good evidence would imply a low $p(e,b)$ value and would indicate high content of the hypothesis. In other words, Popperian corroboration would require a low $p(e,b)$ (Faith and Trueman, 2001, p. 336; Faith, 2006).

Unfortunately, Popper provides no evidence for the decoupling of falsification/falsifiability and corroboration/testability that Faith and Trueman propose. Instead, quite the contrary, Popper consistently links corroboration to falsification. For instance does Popper (1983, p. 231) closely relate degree of corroboration to testability, while claiming that testability can be measured by the content of the theory, which can be measured by the absolute logical improbability of the theory, which is defined as a measure of the class of its falsifiers. As according to Popper corroborability is inversely proportional to absolute logical probability, corroborability equals testability, which equals empirical content, which is measured by the amount of potential falsifiers of the hypothesis and thus its degree of falsifiability (Popper, 1935/1994, pp. 87, 84; Popper,

1983, p. 245). One can unambiguously deduce from Popper's definition of corroborability and falsifiability that in case a hypothesis is not falsifiable it has no corroborability either!

Popper also claims that "only if *e* is the result of genuine or sincere attempts to refute *h* can *e* be regarded as supporting *h*" (Popper, 1983, p. 235), and that "[t]here are two attitudes, two ways of looking at the relations between a theory and experience: one may look for confirmation, or for refutation. (…) *Scientific tests are always attempted refutations"* (Popper, 1983, p. 243). It seems that Faith's demand of a less falsification-centered framework is, to say the least, questionable (Rieppel, 2003). After all, Popper's approach has been called falsificationism and not corroborationism for good reason and Popper himself states: "I may mention here in passing that the idea of the empirical content of a theory, as a measure of the class of its falsifiers, was perhaps the most important *logical* idea of [Logic of Scientific Discovery]. It plays a decisive role in the theory of degrees of testability; of simplicity; of logical probability and improbability; and of corroboration." (Popper, 1983, p. 231).

Another critical point of the "inclusive framework" is its use of goodness-of-fit as evidence for testing phylogenetic tree hypotheses. Evidence in Popperian hypothesis testing represents what Popper called basic statements. Therewith, Popper sets up specific criteria that have to be met to qualify as a basic statement. The most central criterion is that basic statements have to refer to observational events/occurrences (Popper, 1935/1994, p. 68). Moreover, they must be the result of a genuine test (Popper, 1983, p. 254), which means that they have to be testable themselves (Popper, 1935/1994, p. 21). However, as I have already discussed with respect to de Queiroz and Poe's approach, goodness-of-fit of data and cladogram, independent of whether likelihood or parsimony is used, cannot be tested and, thus, cannot serve as source for basic statements. (Faith himself states that the application of goodness-of-fit is not in itself corroboration; Faith, 1992, 2004.). Therefore, although their approach may represent a reasonable method *outside* of Popperian falsificationism, goodness-of-fit does not meet Popper's criteria (for a different argument against the use of goodness-of-fit as evidence in hypothesis testing see Farris et al., 2001).

Final thoughts

In theoretical discussions the congruence test is usually assigned a central function in phylogenetic methodology. I have shown above that the interpretation of the congruence test as a Popperian test has its limitations and requires the matrix (i.e., a set of character and character state hypotheses) to successfully have passed Popperian tests beforehand in order to qualify as basic statements for the congruence test. It has been claimed before that a test of character hypotheses is required that is situated within the character analysis, and similarity has been suggested as a suitable test criterion (Rieppel and Kearney, 2002). However, similarity does not allow for Popperian hypothesis testing as it lacks the necessary deductive link between hypothesis and observational data. Only the criterion of identity as a test of character state hypotheses (Vogt, 2002, 2004a,b) and the criterion of conjunction as a test of character hypotheses (Patterson, 1982; Freudenstein, 2005) may provide a suitable basis for a Popperian testing of character state hypothesis in the former case and character hypotheses in the latter (it goes beyond the scope of this paper to discuss these two tests). Unfortunately, knowing that hypotheses of character states and of characters represent falsifiable phylogenetic hypotheses does not automatically equal having a Popperian rational for the inference of the best tree in the light of the present evidence (especially as the congruence test does not test cladograms but instead sets of hypotheses of apomorphy).

Owing to the fact that there is no criterion for deciding which of the hypotheses of apomorphy represent the falsifying instances in case a set of apomorphy hypotheses fails to pass the congruence test, it would be necessary to quantify the degrees of corroboration of all those apomorphy hypotheses that passed the identity and the conjunction test to be able to quantify the degrees of corroboration that all congruent sets of hypotheses of apomorphy gained by successfully passing the congruence test. This would equal weighting of phylogenetic characters for which, hitherto, no solution has been found on Popperian grounds.

Thus, there is no practicable Popperian approach for the justification of a preference among theoretically possible phylogenetic tree hypotheses due to the fact that no method exists for quantifying the degree of corroboration for cladograms. Moreover, even if one could quantify the degrees of corroboration that character and character state hypotheses gain in the identity and the conjunction test, it is more than questionable whether one can justify a preference for a particular tree among all possible trees on the basis of degrees of corroboration of apomorphy hypotheses. One would have to interpret the degree of corroboration of a character hypothesis as degree of support for or evidence against a given tree hypothesis. If the choice of a best tree would be based on this rational, it would not at all be based on the result of a Popperian test of a tree hypothesis. Strictly speaking, the interpretation of corroborated hypotheses of apomorphy to provide a measure of empirical support for cladograms represents

an inductive step and is incongruent with Popper's approach as cladograms are not falsifiable in principle and, thus, cannot gain corroboration. Consequently, assigning degrees of support to tree hypotheses does not seem to be consistent with Popper's philosophy.

With regard to likelihood or parsimony one has to conclude that asserting, empirical data better supports one specific tree (i.e., the most parsimonious or most likely tree) than all possible alternatives, is not the same as attempting to refute tree hypotheses by contradicting evidence. Therefore, parsimony and likelihood methods pick the tree merely according to their specific goodness-of-fit criterion, which is not based on Popperian falsificationism.

All the conclusions presented here rest on the premise that Popperian falsificationism is the only valid scientific theoretical approach. This is highly questionable and might turn out to be wrong, especially with respect to historical sciences—and phylogenetics represents a historical science. The reason for my doubts refers to the fact that it is in principle impossible to *predict* future observations to test historical hypotheses, which represents the initial idea of Popper's hypothetico-deductive setting—to test necessary predictions (*non-e*) against future observations (*e*: falsification of *h*; *non-e*: corroboration of *h*). This rationale is obviously designed for sciences that use experiments in which they control a set of critical conditions in order to generate the deductively predicted necessary effect in order to test a universal causal hypothesis. Perhaps it is time that phylogeneticists develop their own philosophy, a philosophy of phylogenetics that meets the specific requirements of our scientific field (Rieppel, 2003; Helfenbein and DeSalle, 2005), instead of trying to apply a philosophy like Popper's falsificationism that has been developed for experimental sciences such as physics, which is seeking for universal laws and regularities instead of the reconstruction of particular historical events. However, this problem cannot be dealt with here.

Independent of my concerns regarding the general relevance of falsificationism for phylogenetics, in any science theoretical approach, deduction and hypothesis testing has to take in a central role. Therewith it is important to note that hypothesis testing is not unique to Popperian falsificationism, and not applying falsificationism does not necessarily mean that hypothesis testing does not take in a central methodological function. This holds true for the attempt of reconstructing phylogeny as well. In order to test phylogenetic hypotheses one necessarily has to abstract from the particularity of a specific historical process to be able to apply more general concepts, which, in turn, allow for deduction and for hypothesis testing, the evaluation of putative supporting evidence, and the elimination of misleading evidence.

The result of this study might be disappointing for all those who want to rest phylogenetic methodology on the seemingly consistent and seemingly invulnerable methodological position of Popperian falsificationism. However, the foundations of falsificationism have been questioned and criticized before and the limits of its practical application are well known (e.g., Salmon, 1967, 1968, 1998; Lakatos, 1968, 1970; Kuhn, 1970; Putnam, 1974; Grunbaum, 1976; Mackie, 1985; Sober, 1988, 2000; Howson and Urbach, 1989; Earman and Salmon, 1992; McGuire, 1992; Stamos, 1996; Andersson, 1998; Schurz, 1998; Franklin, 2001; Spohn, 2001). It seems to be very difficult, if not impossible in principle, to apply falsificationism consistently within the historical sciences such as phylogenetics. Thus, there exists a need for a sound epistemological foundation for phylogenetics independent of Popperian falsificationism. In this respect, phylogenetics proves to be a challenging scientific enterprise. This is not at all surprising—no one should expect that the reconstruction of single transformation events that took place millions of years in the past within the ancestral lineages of representatives of our recent species is an easy task that involves no methodological problems. In phylogenetics we cannot rerun evolution—we cannot test statements of universal evolutionary laws. Instead, we want to give explanations for particulars that are the result of a specific sequence of causal events that we call phylogeny. The epistemic situation we come across in phylogenetics constitutes the uniqueness of phylogenetic research within the biological sciences but is also responsible for the specific methodological problems and the inevitable necessity of making assumptions whose testability is limited and who nonetheless affect the outcome of phylogenetic analyses significantly.

Acknowledgments

I thank Gonzalo Giribet, Thomas Bartolomaeus and Christoph Bleidorn for reading and criticizing the earlier drafts of this manuscript. Furthermore I thank Olivier Rieppel and Andy Brower for reviewing the manuscript and giving me valuable comments and suggestions. It goes without saying, however, that I am solely responsible for all the arguments and statements in this paper. This study was supported by the Deutsche Forschungsgemeinschaft DFG (BA 1520/4-2 and VO 1244/1-1).

References

Andersson, G., 1998. Basisprobleme. In: Keuth, H. (Ed.), Klassiker Auslegen: Karl Popper – Logik der Forschung. Akademie Verlag, Berlin, pp. 145–164.

Bock, W.J., 1969. Nonvalidity of the 'phylogenetic fallacy'. Syst. Zool. 18, 111–115.
Bock, W.J., 1973. Philosophical foundations of classical evolutionary classification. Syst. Zool. 22, 375–392.
Bryant, H.N., 1992. The role of permutation tail probability tests in phylogenetic systematics. Syst. Biol. 41, 258–263.
Cracraft, J., 1978. Science, philosophy, and systematics. Syst. Zool. 27, 213–216.
Cracraft, J., Helm-Bychowski, K., 1991. Parsimony and phylogenetic inference using DNA sequences: Some methodological strategies. In: Miyamoto, M.M., Cracraft, J. (Eds.), Phylogenetic Analysis of DNA Sequences. Oxford University Press, New York, pp. 184–220.
Duhem, P., 1906. La théorie physique, son objet et sa structure. Marcel, Paris. In: Wiener, P.P. (Ed.), The Aim and Structure of Physical Theory, Princetown University Press, Trans. Princetown, NJ, 1954.
Earman, J., Salmon, W.C., 1992. The confirmation of scientific hypotheses. In: Salmon, M.H., Earman, J., Glymour, C., Lennox, J.G., Machamer, P., McGuire, J.E., Norton, J.E., Salmon, W.C., Schaffner, K.F. (Eds.), Introduction to the Philosophy of Science. Prentice Hall, Englewood Cliffs, NJ, pp. 42–103.
Engelmann, G.F., Wiley, E.O., 1977. The place of ancestor-descendant relationships in phylogeny reconstruction. Syst. Zool. 26, 1–11.
Faith, D.P., 1991. Randomization tests for monophyly and non-monophyly. In: Hennig X. Willi Hennig Society, Toronto, Canada.
Faith, D.P., 1992. On corroboration: a reply to Carpenter. Cladistics 8, 265–273.
Faith, D.P., 1999. Error and the growth of experimental knowledge [Review]. Syst. Biol. 48, 675–679.
Faith, D.P., 2004. From species to supertrees: Popperian corroboration and some current controversies in systematics. Aust. Syst. Bot. 17, 1–16.
Faith, D.P., 2006. Science and philosophy for molecular systematics: Which is the cart and which is the horse? Mol. Phylogenet. Evol, 38, 553–557.
Faith, D.P., Cranston, P.S., 1992. Probability, parsimony, and Popper. Syst. Biol. 41, 252–257.
Faith, D.P., Trueman, J.W.H., 2001. Towards an inclusive philosophy for phylogenetic inference. Syst. Biol. 50, 331–350.
Farris, J.S., 1983. The logical basis of phylogenetic analysis. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in Cladistics, Vol. 2. Columbia University Press, New York, pp. 7–36.
Farris, J.S., 1986. On the boundaries of phylogenetic systematics. Cladistics 2, 14–27.
Farris, J.S., 1995. Conjectures and refutations. Cladistics 11, 105–118.
Farris, J.S., 2000. Corroboration versus 'Strongest Evidence'. Cladistics 16, 385–393.
Farris, J.S., Kluge, A.G., Carpenter, J.M., 2001. Popper and likelihood versus "Popper*". Syst. Biol. 50, 438–444.
Franklin, J., 2001. Resurrecting logical probability. Erkenntnis 55, 277–305.
Freudenstein, J.V., 2005. Characters, states and homology. Syst. Biol. 54, 965–973.
Grant, T., Kluge, A.G., 2003. Data exploration in phylogenetic inference: scientific, heuristic, or neither. Cladistics 19, 379–418.
Grunbaum, A., 1976. Is the method of bold conjectures and attempted refutations justifiably the method of science. Br. J. Philos. Sci. 27, 105–136.
Helfenbein, G.K., DeSalle, R., 2005. Falsifications and corroborations: Karl Popper's influence on systematics. Mol. Phylogenet. Evol, 35, 271–280.
Howson, C., Urbach, P., 1989. Scientific Reasoning: The Bayesian Approach. Open Court Publishing Co., La Salle, IL.
Hull, D.L., 1980. Karl Popper and Plato's metaphor. In: Platnick, N., Funk, V.A. (Eds.), Advances in Cladistics, Vol. 1. Columbia University Press, New York, pp. 177–189.
Hull, D.L., 1983. Karl Popper and Plato's Metaphor. In: Platnick, N.I., Funk, V.A. (Eds.), Advances in Cladistics II. Columbia University Press, New York, pp. 177–189.
Hume, D., 1748/1993. An Enquiry Concerning Human Understanding. Hacking Publishing Co., Indianapolis, IN.
Kearney, M., Rieppel, O., 2006. Rejecting 'the given' in systematics. Cladistics 22, 369–377.
Kitts, D.B., 1977. Karl Popper, verifiability, and systematic zoology. Syst. Zool. 26, 185–194.
Kluge, A.G., 1997a. Testability and the refutation and corroboration of cladistic hypotheses. Cladistics 13, 81–96.
Kluge, A.G., 1997b. Sophisticated falsification and research cycles: Consequences for differential character weighting in phylogenetic systematics. Zool. Scr. 26, 349–360.
Kluge, A.G., 1999. The science of phylogenetic systematics: explanation, prediction, and test. Cladistics 15, 429–436.
Kluge, A.G., 2001a. Philosophical conjectures and their refutation. Syst. Biol. 50, 322–330.
Kluge, A.G., 2001b. Parsimony with and without scientific justification. Cladistics 17, 199–210.
Kluge, A.G., 2002. Distinguishing 'or' from 'and' and the case for historical identification. Cladistics 18, 585–593.
Kluge, A.G., 2003. The repugnant and the mature in phylogenetic inference: atemporal similarity and historical identity. Cladistics 19, 356–368.
Kuhn, T.S., 1970. Logic of discovery or psychology of research? In: Lakatos, I., Musgrave, A. (Eds.), Criticism and the Growth of Knowledge. Cambridge University Press, Cambridge, pp. 1–23.
Lakatos, I., 1968. Changes in the problem of inductive logic. In: Lakatos, I., eds. The problem of inductive logic. North-Holland, Amsterdam, pp. 315–417.
Lakatos, I., 1970. Falsification and the methodology of scientific research programmes. In: Lakatos, I., Musgrave, A. (Eds.), Criticism and the Growth of Knowledge. Cambridge University Press, Cambridge, pp. 91–196.
Mackie, J.L., 1985. Logic and Knowledge. Selected Papers, Vol. I. Clarendon Press, Oxford.
McGuire, J.E., 1992. Scientific change: perspectives and proposals. In: Salmon, M.H., Earman, J., Glymour, C., Lennox, J.G., Machamer, P., McGuire, J.E., Norton, J.E., Salmon, W.C., Schaffner, K.F. (Eds.), Introduction to the Philosophy of Science. Prentice Hall, Englewood Cliffs, NJ, pp. 132–178.
Niiniluoto, I., 1995. Covering law model. In: R. Audi (Ed.), The Cambridge Dictionary of Philosophy. Cambridge University Press, New York, pp. 190–192.
O'Hara, R.J., 1997. Population thinking and tree thinking in systematics. Zool. Scr. 26, 323–329.
Patterson, C., 1982. Morphological characters and homology. In: Joysey, K.A., Friday, A.E. (Eds.), Problems of Phylogenetic Reconstruction. Academic Press, London, pp. 21–74.
Patterson, C., 1988. Homology in classical and molecular biology. Mol. Biol. Evol. 5, 603–625.
de Pinna, M.C.C., 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics 7, 367–394.
Platnick, N.I., Gaffney, E.S., 1977. Systematics: a Popperian perspective. Syst. Zool. 26, 360–365.
Platnick, N.I., Gaffney, E.S., 1978. Systematics and the Popperian paradigm. Syst. Zool. 27, 381–388.
Popper, K.R., 1935/1994. Logik der Forschung, 10. Auflage. J. C. B. Mohr (Paul Siebeck), Tübingen.
Popper, K.R., 1983. Realism and the Aim of Science; from the Postscript to the Logic of Scientific Discovery. Routledge, London.
Putnam, H., 1974. The 'Corroboration' of Theories. In: Schlipp, P.A. (Ed.). The Philosophy of Karl Popper, Book I. Open Court, La Salle, IL, pp. 221–240.
de Queiroz, K., Poe, S., 2001. Philosophy and phylogenetic inference: a comparison of likelihood and parsimony methods in the context

of Karl Popper's writings on corroboration. Syst. Biol. 50, 305–321.
de Queiroz, K., Poe, S., 2003. Failed refutations: further comments on parsimony and likelihood methods and their relationship to Popper's degree of corroboration. Syst. Biol. 52, 352–367.
Quine, W.V.O., 1951. Two dogmas of empirism. The Philos. Rev. 60, 20–43.
Rieppel, O., 2003. Popper and systematics. Syst. Biol. 52, 259–271.
Rieppel, O., 2005. The philosophy of total evidence and its relevance for phylogenetic inference. Papéis Avulsos Zool. 45, 77–89.
Rieppel, O., Kearney, M., 2002. Similarity. Biol. J. Linn. Soc. Lond. 75, 59–82.
Salmon, W., 1967. The Foundations of Scientific Inference. University of Pittsburgh Press, Pittsburgh, PA.
Salmon, W.C., 1968. The Justification of Inductive Rules of Inference. In: Lakatos, I. (Ed.). The Problem of Inductive Logic. North-Holland, Amsterdam, pp. 24–43.
Salmon, W.C., 1998. Rational prediction. In: Curd, M., Cover, J.A. (Eds.), Philosophy of Science: the Central Issues. W. W. Norton Co., Inc., New York, pp. 433–444.
Schroeder-Heister, P., 1998. Wahrscheinlichkeit. In: Klassiker Auslegen: Karl Popper – Logik der Forschung. Akademie Verlag, Berlin, pp. 185–213.
Schurz, G., 1998. Das Problem der Induktion. In: Keuth, H. (Ed.), Klassiker Auslegen: Karl Popper – Logik der Forschung. Akademie Verlag, Berlin, pp. 25–40.
Settle, T., 1979. Popper on 'When is a science not a science?'. Syst. Zool. 28, 521–529.
Siddall, M.E., Kluge, A.G., 1997. Probabilism and phylogenetic inference. Cladistics 13, 313–336.
Sober, E., 1983. Parsimony in systematics: Philosophical issues. Annu. Rev. Ecol. Systematics 14, 335–357.
Sober, E., 1988. Reconstructing the Past. Parsimony, Evolution, and Inference. MIT Press, Cambridge, MA.
Sober, E., 2000. Philosophy of Biology, 2nd edn. Westview Press, Boulder, CO.
Spohn, W., 2001. Vier Begründungsbegriffe. In: Grundmann, T. (Ed.), Erkenntnistheorie – Positionen zwischen Tradition und Gegenwart. Mentis, Paderborn, pp. 33–52.
Stamos, D.N., 1996. Popper, falsifiability, and evolutionary biology. Biol. Philos. 11, 161–191.
Vogt, L., 2002. Testing and weighting characters. Organisms, Diversity Evol. 2, 319–333.
Vogt, L., 2004a. Signs and phylogeny: a semiotic approach to systematics. Semiotica 149, 125–159.
Vogt, L., 2004b. Zeichenprozesse in der Phylogenie und der Phylogenetik – eine wechselseitige Erhellung. In: Schult, J. (Ed.), Biosemiotik—Praktische Anwendung und Konsequenzen für Einzelwissenschaften. Studien zur Theorie der Biologie Band, 6. Verlag für Wissenschaft und Bildung (VWB), Berlin (O. Breidbach & M. Weingarten), pp. 45–78.
Wagner, G.P., 1986. The systems approach: an interface between development and population genetic aspects of evolution. In: Raup, D.M., Jablonski, D. (Eds.), Patterns and Processes in the History of Life. Springer, Berlin, pp. 149–165.
Wiley, E.O., 1975. Karl R. Popper, systematics, and classification: a reply to Walter Bock and other evolutionary systematists. Syst. Zool. 24, 233–243.
Wilson, E.O., 1965. A consistency test for phylogenies based on contemporaneous species. Syst. Zool. 14, 214–220.